AFRL-HE-AZ-TR-2002-0207

# UNITED STATES AIR FORCE RESEARCH LABORATORY

## ASSESSMENT OF HUMAN INTERACTION WITH VIRTUAL ENVIRONMENT TRAINING TECHNOLOGY


**CELESTINE A. NTUEN**
**The Institute for Human Machine Studies**
**Department of Industrial and Systems Engineering**

**S. YOON**
**Department of Computer Science**

**North Carolina A&T State University**
**1601 East Market Street**
**Greensboro, NC 27411**


**October 2002**

Approved for public release; distribution is unlimited.


**AIR FORCE MATERIEL COMMAND**
**AIR FORCE RESEARCH LABORATORY**
**Human Effectiveness Directorate**
**Warfighter Training Research Division**
**6030 South Kent Street**
**Mesa AZ 85212-6061**

## NOTICES

**The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of this organization or the US Government.**

**Publication of this report does not constitute approval or disapproval of the ideas or findings. It is published in the interest of STINFO exchange.**

**Using Government drawings, specifications, or other data included in this document for any purpose other than Government-related procurement does not in any way obligate the US Government. The fact that the Government formulated or supplied the drawings, specifications, or other data, does not license the holder or any other person or corporation, or convey any rights or permission to manufacture, use, or sell any patented invention that may relate to them.**

**The Office of Public Affairs has reviewed this paper, and it is releasable to the National Technical Information Service, where it will be available to the general public, including foreign nationals.**

**This report has been reviewed and is approved for publication.**

| | |
|---|---|
| **PETER CRANE**<br>**Project Scientist** | **DEE H. ANDREWS**<br>**Technical Advisor** |

**CURTIS J. PAPKE, Colonel, USAF**
**Chief, Warfighter Training Research Division**

Copies of this report may be requested from:

Defense Technical Information Center
8725 John J. Kingman Road, Suite 0944
Ft. Belvoir, VA 22060-6218
http:// stinet.dtic.mil

# REPORT DOCUMENTATION PAGE

*Form Approved*
*OMB No. 0704-0188*



**1. REPORT DATE** *(DD-MM-YYYY)*
October 2002

**2. REPORT TYPE**
Final

**3. DATES COVERED** *(From - To)*
Nov 99 to Jun 02

**4. TITLE AND SUBTITLE**
Assessment of Human Interaction with Virtual Environment Training Technology

**5a. CONTRACT NUMBER**
F41624-00-1-0001

**5b. GRANT NUMBER**

**5c. PROGRAM ELEMENT NUMBER**
62205F

**6. AUTHOR(S)**
Celestine A. Ntuen
S. Yoon

**5d. PROJECT NUMBER**
1123

**5e. TASK NUMBER**
B1

**5f. WORK UNIT NUMBER**
17

**7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)**
North Carolina A&T State University
1601 East Market St.
Greensboro, NC 27411

**8. PERFORMING ORGANIZATION REPORT NUMBER**

**9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)**
Air Force Research Laboratory
Human Effectiveness Directorate
Warfighter Training Research Division
6030 South Kent Street
Mesa AZ 85212-6061

**10. SPONSOR/MONITOR'S ACRONYM(S)**
AFRL

**11. SPONSOR/MONITOR'S REPORT NUMBER(S)**
AFRL-HE-AZ-TR-2002

**12. DISTRIBUTION / AVAILABILITY STATEMENT**
Approved for public release; distribution is unlimited.

**13. SUPPLEMENTARY NOTES**
Air Force Research Laboratory Technical Monitor: Dr Peter Crane, AFRL/HEAS, 480.988.6561 x-287, DSN 474-6287.

**14. ABSTRACT**
This research investigated the evidence of performance improvement of piloting skills while using an immersive virtual environment (IVE) versus a nonimmersive virtual environment (NIVE) to train instrument pilot skills. The general hypothesis tested was whether there is equal improvement for people trained under IVE and NIVE. Subjects were tested in IVE and NIVE flight scenarios using three flying tasks—normal crosswind approach and landing (NCAL), go-around (GA), and constant speed during climbing and descending (CSCD). Data were analyzed for two measures – errors and error rate, for four dependent variables: altitude control, heading control, airspeed control, and vertical airspeed control. Overall, results failed to demonstrate enhanced training effectiveness for an immersive VR training environment compared to a desktop (nonimmersive) environment. These results indicate that the cost tradeoff between the uses of IVE over NIVE are task dependent and influenced by the fidelity of training environments. The results obtained from the current experiment do justify some potential cost-saving advantage of IVE over NIVE on selected task. For example, NIVE seems to provide training advantages on error rate reduction on control of vertical airspeed and altitude under NCAL tasks. Similarly, IVE seems to offer training advantages of error rate reduction on airspeed control and heading control under NCAL, and heading and vertical airspeed controls under GA tasks. However, the fact that either IVE or NIVE provides an increase in piloting task performance in some tasks needs to be considered in any training investment decision.

**15. SUBJECT TERMS**
Human interaction; Performance; Piloting skills; Immersive virtual environment; IVE; Nonimmersive virtual environment; NIVE; Training effectiveness; Virtual Environment Training Technology; Virtual environments;

| 16. SECURITY CLASSIFICATION OF: | | | 17. LIMITATION OF ABSTRACT | 18. NUMBER OF PAGES | 19a. NAME OF RESPONSIBLE PERSON |
|---|---|---|---|---|---|
| | | | | | Ms Liz Casey |
| **a. REPORT** | **b. ABSTRACT** | **c. THIS PAGE** | | | **19b. TELEPHONE NUMBER** *(include area code)* |
| UNCLASSIFIED | UNCLASSIFIED | UNCLASSIFIED | UNLIMITED | 50 | 480.988.6561 x-188 DSN 474-6188 |

**Standard Form 298 (Rev. 8-98)**
**Prescribed by ANSI Std. Z39.18**

CONTENTS

## LIST OF FIGURES

**FIGURE** **PAGE**

## LIST OF TABLES

**TABLE** **PAGE**

# PREFACE

This report documents the results of a project conducted by The Institute for Human-Machine Studies at North Carolina A&T State University under U.S. Air Force Grant No. F41624-00-1-0001, Workunit No. 1123B117, Human Interaction with Virtual Environment Training Technology. Dr. Peter Crane is the Laboratory Contract Monitor.

The goal of the research was to investigate if there are enhanced training benefits resulting from using an immersive virtual environment (IVE) versus a nonimmersive virtual environment (NIVE) in piloting training tasks.

During the course of the research, two graduate students performed most of the work. Mr. Robert Halpin was responsible for developing the software system for both IVE and non-NIVE. Mr. Halpin is a graduate student in the Department of Computer Science. Mr. Kaize Adams, a graduate student in the Department of Industrial and Systems Engineering was responsible for human factors experiments. A part of this report serves as a thesis report for Mr. Adams' requirement for a master's degree in Industrial and Systems Engineering. Both these graduate students provided the technical assistance required for the project completion.

## ACRONYMS

AE: average error (deviations from standard performance measures), an absolute measure of performance
AER: average error (error values per time unit), standard or established metric of performance for minimum skill learning proficiency as used by flight standard instructors
ALT: altitude (ft), flight task variable
APC: acceptable performance criterion
APM: absolute performance measure
AS: airspeed (knots), flight task variable
ATC: air traffic control, ground control resource
CTA: cognitive task analysis, a method for acquiring cognitive knowledge related to work analysis
COTS: commercial-off-the-shelf, a suite of software product available in commercial release
CSCD: constant airspeed during climbing and descending, performed by pilots to test climbing and descending at given speeds
EFI: expert flight inspector
FAA: Federal Aviation Administration, a government body responsible for commercial and civil aviation
FOR: field –of- regard, a point in a display space
FOV: filed-of-view, a defined angle of vision with respect to the environment
GA: go-around task, performed by pilots to test constant air or ground speed control
HCI: human-computer interface, the study of human-machine communication and information display
HMD: head-mounted display, a hardware device used in to render information in a virtual environment
ILS: Instrument landing system
IVE: immersive virtual environment, a typical mode of enactive interaction in which the operator's perception is tightly coupled with the environment
KA: knowledge acquisition, a term used to conduct cognitive task analysis
NCAL: normal crosswind approach and landing, a flight task that deals with approach and landing on a runway with a moderate wind
NIVE: non immersive virtual environment, a typical active interaction in which the operator uses environment-perception decoupled display to perform tasks; mostly with desktop computers or large screen displays
PC: personal computer
SAS: statistical analysis software
SDK: software development kit in Microsoft Flight Simulation 2000
SME: Subject-matter experts, people with experience often used in walkthrough studies in CTA
STE: skill training enhancement
VE: virtual environment, usually a workspace with computer enhanced display of reality rendered in three-dimensions (3-D)
VETS: virtual reality training system, a software developed for this project
VR: virtual reality, a virtual environment in which the operator is perceptually coupled with the task by wearing display enhanced systems such as stereoscopic or binocular displays, 3-D sounds, and other enabled display technologies

# ASSESSMENT OF HUMAN INTERACTION WITH VIRTUAL ENVIRONMENT TRAINING TECHNOLOGY

# 1. INTRODUCTION

## 1.1 Background

Skill is defined as the learned ability of associating an optimal action with the task process state or its characteristics (Bilodeau & Bilodeau, 1961; Fitts, 1967). According to Rasmussen (1986), the skill level is "where automated routines are based on subconscious time-space manipulations of objects or symbols in familiar scenery (pp.113)." Learning has the traditional view that implies change in behavior through acquisition of some skill (Gagne, 1962). Such a behavior change may be simple adaptation to a new situation, or a gradual shift in the level of "expertise" resulting in rich knowledge content.

Training for pilot skill acquisition has been and will continue to be of great concern as new methods and technologies unfold. The two most important piloting tasks often considered for training priorities are aviation and navigation tasks. To aviate refers to the fact that pilots must control the aircraft's path. They are responsible for controlling their aircraft along three-dimensional (3-D) and three angular axes. The navigation of the aircraft means that the pilots must maneuver their aircraft from one location to another in time and space (Caro, 1998; Connolly, Blackwell, & Lester, 1989).

In a very important assertion, Dennis and Harris (1998) note that flight training is expensive, and as a result, methods to reduce the cost of training are constantly being sought. Among the several methods of pilot training available (Koonce & Bramble, 1998; Lintern, Thomely-Yates, Nelson & Roscoe, 1987; Povenmire & Roscoe, 1971), virtual reality (VR) is being investigated as a training environment for developing realistic training systems. With VR environments, flying tasks can be made more realistic with spatial objects showing landmarks, terrain, weather, and other situation aids in a virtual environment, yet powerful enough to replicate the real flying tasks. For example, virtual task shells and program tools allow piloting task scripts to be represented with multimedia tools (texts and video), as well as allowing these tools to be embedded into a virtual environment. This allows for improvisation of realistic task knowledge within the training software (Augusteijn, Broome, Kolbe, & Ewell, 1992; Chambers & Nagel, 1985).

Assessing a pilot's performance in aviation and navigation tasks has continued to elude pilot trainers who rely on high-fidelity simulation environments to measure training and skill acquisitions of student pilots during flight training. With the availability of VR training systems, the issue of performance comparison between current computer-based, motion-driven, high-fidelity flight simulators has become necessary. Such a comparative study is needed to determine the tradeoff in deciding which training mode is most efficient with respect to learning flight skill acquisition and reduction in piloting errors (Dion, Smith, & Dismukes, 1996; Koonce, 1984).

This particular experiment will compare student performance in selected flying tasks to published standards. The results can be used to assess the impact of different simulator configurations on training effectiveness

Virtual reality training environments have rich domains to represent task knowledge with near realism and fidelity (De Keysers, 1987; Hirtle & Hudson, 1991). This characteristic makes it attractive to study the effect of pilot's attention between the physical world and virtual tasks (Psotka, 1995; Regian & Shebilske, 1992). On the other hand, the existing computer-based training systems that do not provide direct conscious immersion in the environment can only induce artificial experience with animation and rich graphical representation of tasks. In this kind of training system, the user and the simulated tasks are isolated and interactions are inactive. To truly understand the potential cost saving between the two computer-based training systems, we have developed a Virtual Environment Training System (VETS) that has both the features of immersive and nonimmersive environments.

## 1.2 Learning versus Training

Training is a construct used for developing specific skills required to perform specific tasks. The effectiveness of training is realized through learning (Bell & Waag, 1998). Figure 1 shows the relationship between concept learning and skill training for specific tasks.

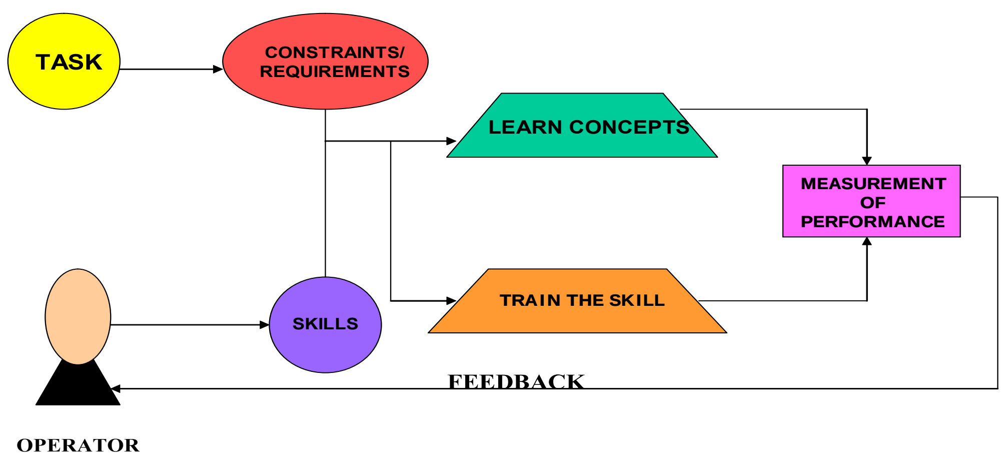


Figure 1. Relationship between Concept Learning and Skill Training for Specific Tasks.

Improving competence in procedure, comprehension, and operation has been the emphasis in aircraft piloting trainability factors (Johnston & Maurino, 1990; Lintern, 1980). There are many other reports that show the utility of virtual reality in holistic training. Jaeger (1998) notes that virtual reality environments reach sensory input that lead to more immersive experience making it possible to train a person by using single or multiple sensory modalities. However, performance gains from training with VR varies across individuals and tasks (Piantanida, Boman, & Gille, 1995; McCreary & Williges, 1998). Most importantly, the focus of VR training has been on navigation and spatial awareness tasks. These tasks involve subjects to learn and memorize landmarks and spatially distributed geometric networks. An example is a military pilot flying in a rectangular pattern to protect a target. The pilot is trained to memorize the rectangular maze or configuration. An example of a learning task is maneuvering. Maneuvering has navigation as its subset. Maneuvering involves an intended and controlled variation from a straight-and-level flight path in the operation of an airplane. In this research, significant parts of the tasks are considered to be maneuvering.

## 1.3 Pilot Skill Learning and Virtual Reality Environment

New methods of training pilots are always being sought to reduce the high costs for training. In recent years virtual environments have been used to aid in the training process of various tasks; for example, learning relative directions between landmarks and navigating through virtual buildings.

A virtual environment is usually a computer-generated, 3-D environment in which a person is immersed. These environments can be immersive through the use of head-mounted displays or nonimmersive through the use of desktop monitors. In VR systems, immersive experience is provided by body-worn, visually coupled displays with view stereoscopic or biocular images, 3-D sounds, and with environmental images rendered in a manner similar to the real 3-D world (Pimental & Teixera, 1995).

In recent years, psychologists and scientists have been researching the advantages and differences of learning in VEs. Their studies have dealt with various issues such as training and learning. For example, studies conducted by Albert, Rensink, and Beusmans (1999); Harper, (1991), Patrick, et al. (2000), Ruddle, Payne, and Jones (1999), and Rose and Attree ( 2000) confirm the effectiveness of using VEs for training.

## 1.4 Objective and Scope

### 1.4.1 Objectives

This experiment will investigate whether there is training improvement in flight performance while using an immersive virtual environment (IVE) and a nonimmersive virtual environment (NIVE). Both IVE and NIVE task scenarios were developed using Microsoft's "Flight Simulator 2000: Professional Edition" software. The flight simulation software measured and recorded performance of the participants. Given the expense of helmet-mounted displays (HMDs), limited resolution, the requirement for head tracking, and additional software development, it would be useful to have data relating to the unique advantages of immersive environments. The main hypothesis investigated can be posed as: *Does piloting task performance improve more in IVE than NIVE?* By investigating this general hypothesis, we seek to answer the following questions:

Question 1: Are there statistically significant differences in average task performance error between subjects trained under IVE and NIVE?

Question 2: Are there statistically significant differences in task performance error rate between subjects trained under IVE and NIVE?

Question 3: Does IVE provide better pilot skill training enhancement than NIVE?

### 1.4.2 Scope

Data on altitude, airspeed, heading, and vertical speed were collected from subjects during simulated flight experiments using the VETS software designed for this project. The main performance data are error and error rate obtained from selected flight scenarios in the simulation experiments. An average error is used to measure absolute performance, and average error rate is used as a measure of acceptable operation performance for minimum skill learning as used by flight standard examinations.

The participants used the VETS environment to perform three piloting tasks considered to be most complex to learn according to data obtained from subject-matter experts (SMEs). The tasks were selected from the Federal Aviation Administration (FAA) Practical Test Standard FAA-S-8081-14. The tasks were; maintaining constant airspeed during climbing and descending, go-around, and normal crosswind approach and landing.

# IMMERSIVE AND NONIMMERSIVE VIRTUAL ENVIRONMENTS FOR SKILL TRAINING

## 2.1 Immersive Virtual Environments

The typical IVE sensory input to the human from the external world is, ideally and often wholly, provided by the computer-generated displays (Slater & Usoh, 1993). This sense of immersion is a very important factor relevant to training performance. Aukstakalnis and Blatner (1992) describe the condition of immersion as follows:

> Being immersed means being surrounded by something; everywhere you look, it's there. To create a sense of immersion in a virtual environment, we must be able to surround ourselves with various stimuli in a manner that makes sense and that follows roles similar to those of the real world. That is, when you turn your head to the left, you see the objects to the left of you. When you look forward, you get closer to the objects in front of you. These are elementary features of our sense of being immersed in an environment; and when you're in a virtual environment, you expect the same results. (p. 27)

The typical mode of immersion in an IVE is via a head-tracked display or HMD. Wherever the participant looks, the computer renders the appropriate view to be seen in real or near-real time. Sometimes, 3-D sound is provided through earphones (Wenzel, 1992). Sound appears to the user with those of the virtual sound source. Interaction with the virtual environment (VE) may be limited to locomotion through it, or may include locomotion plus interaction with virtual objects, such as pushing virtual buttons, opening virtual doors, or moving and grasping virtual objects. Some performance problems with HMD includes, but is not limited to field-of-view and total field-of-regard (Gallimore, Brannon, & Patterson, 1998 ), and display resolution and total field-of-regard (Naish & Miller, 1980).

## 2.2 Nonimmersive Virtual Environments

NIVEs are represented by desktop displays. NIVEs allow the participant to interact with a VE and feel a sense of immersion but without the use of an HMD (Ruddle, et al., 1999) note that people typically use abstract interfaces (e.g., mouse, keyboard, joystick, or a spaceball) to control their translationary movements and changes of direction with desktop displays. Further, when using desktop displays, people receive feedback on their movements from visual changes in the displayed scene and the motor actions of their fingers on the interface devices. A common problem with NIVE is limited peripheral vision resulting in irregular eye scanpath and increased information search (Fisher, 1979); this often lead to increase in workload--an attribute responsible for degradation in performance (Yeh & Wickens, 1997).

## 2.3 Differences between IVE and NIVE

Ruddle et al. (1999) and Patrick, et al. (2000) presented the differences between IVEs and NIVEs. Both studies found that there was no significant difference between the two types of training environments with respect to spatial knowledge acquisition tasks. In NIVE, people receive feedback on their movements from visual changes in the displayed scene and the motor actions of their fingers on the interface devices. By contrast, the visual feedback that people receive when using IVE displays is supplemented by vestibular (equilibrium) and kinesthetic (body position) from their changes in direction. Ruddle (as cited in Presson & Montello, 1994; Rieser, 1989; Reiser, Lockman, & Pick, 1980) showed that the effect of this additional feedback on the user's ability to navigate is not known, but data from some real-world studies suggest that feedback helps users to develop spatial knowledge and the physical changes of direction are more important than physical translationary movements for the development of that knowledge.

Patrick et al. (2000) observed that an IVE allows the user to have increased peripheral vision and capability to freely look around the virtual environment. In the NIVE, users tend not to look around because their peripheral vision of the VE is not as large. The advantage of the IVE helps in perception of the VE but a user's sense of presence may also vary between "being inside" immersive VEs and "looking into" desktop VEs, but the effect of presence on the user's ability to navigate in VEs remains to be investigated (Ruddle et al., 1999).

As far as task presentation and comprehension, it was found by Ruddle et al. (1999) that of the participants who navigated the virtual buildings in their study, 12% performed faster and attained more accuracy when using the HMD. This occurrence was attributed to the effect of the HMD, which provided an interface in which changes in view direction were natural (i.e., head and body movements) and required less effort (e.g., quick glances, rather than holding down a mouse button).

## 2.4 Skill Learning in Virtual Environments

Albert et al. (1999) studied the learning of relative directions between landmarks in desktop virtual environments or NIVEs. The results of their study found that subjects learned relative directions between landmarks equally well when scenes were presented in either a sequential or random order. Furthermore, viewing a configuration of landmarks in a desktop virtual environment from multiple perspectives produced a viewpoint dependent representation in memory.

A study by Rose and Attree (2000) measured and evaluated what is transferred from training a simple sensorimotor task in a virtual environment to the real world. It was found that the only significant difference was real task performance after training in a VE was less affected by concurrently performed interference tasks than was real task performance after training on the real task. Further, it was observed that virtual training resulted in equivalent or better to real-world performance than real training in the simple sensorimotor task.

Harper (1991) compared the feasibility of utilizing relatively inexpensive personal computers to teach instrument flying skills or pilot skills. An experiment was designed to compare the transfer of training between the FAA-approved ATC-710, a cab simulator, and Microsoft FlightSim™ 4.0. The task was a controlled fly-off between two groups of participants in an actual aircraft. The groups were named PC groups and ATC groups, respectively. The following results were found:

(i) No statistical significant difference was found between the two simulator groups. Harper stated that it would appear that personal computer technology might be sufficiently mature to be used as cost-effective instrument trainers by general aviation pilots.
(ii) The Microsoft FlightSim™ 4.0 system was particularly more sensitive to pitch control and lacked realism in yaw control as well.
(iii) The magnitudes of rate of descent and rate of climb were unrealistically high. In spite of these differences, all of the PC group participants were successfully trained to fly the flight test profile, and their performance on the flight test profile during the final simulator session was similar to the ATC group.

The reason for the success of the PC group in the study was the task realism that represented the instrument procedures used to control aircrafts.

# 3. USING COGNITIVE TASK ANALYSIS FOR VETS SOFTWARE DESIGN

## 3.1 Application of Cognitive Task Analysis for Developing Pilot Training Tasks

Modern aircraft systems impose multiple, concurrent task demands on the operator. Therefore, for the development of effective training systems, the tasks often engaged by pilots must be well understood and represented in the computer-based training system. The pilot must interact with the automation. In most cases, the necessary evil is cognitive workload. Cognitive workload refers to the portion of operator information processing capacity or resources that are actually required to meet system demands (Chou, Madhavan, & Funk, 1996).

An assessment of the human workload associated with such multitask environments has, therefore, become an important issue in system test and evaluation. This itself is a requirement for training. For example, during task processing, cognitive overload may occur when information processing demand is too high. Similarly, a cognitive underload occurs when the demand is very low (Helmreich, 1984; Kelly, 1988). In both cases, a measure of cognitive workload can be used to determine system efficiency and performance, including a decision on function allocation among humans and machines.

## 3.2 Using CTA for Knowledge Acquisition in Training System Design

Knowledge acquisition is the process for gathering data, information, and knowledge about a task. Cognitive task analysis as discussed before is just one of the several tools to accomplish this. Knowledge acquisition (KA) in practice is itself a complex task, time consuming, and often unsuccessful (Seamester, Redding, & Kaempf, 1997). The limitations are subject to the nature of the task and the availability of SMEs. For example, some KA problems may be attributed to one or several of the following reasons:

(1) Inability of SMEs to articulate their problem-solving skills;
(2) Inability of the knowledge engineers to elicit the appropriate knowledge from the SMEs;
(3) Inconsistency in the SMEs' description of their problem-solving strategies;
(4) Incompleteness in the description of their problem-solving strategies;
(5) Inability of the knowledge engineers to understand the SMEs description of the problem-solving strategies.

The KA strategy adopted for this research used SMEs from Guildford Technical Community College (GTCC). The SME group consisted of flight instructors, a military fighter pilot, and pilots with a commercial flying license. The KA tasks consist of:

**(i) Review of flight lesson plans**: We reviewed flight lesson plans used by GTCC instructors. With this review, we collected training data on: (1) the aspects of flight instruction that is most difficult; (2) the basic intervention strategies used by instructors to improve flight training; and (3) measures of flight training competency and proficiency.

**(ii) Preflight briefing**: We reviewed videotapes of all aspects of preflight briefing used by the instructors. This allowed us to format the instructional strategies for conducting experiments with the virtual reality training system.

**(iii) Other Phases of Flight Tasks:** Other information gathered consists of:
(1) Flight Configuration Data: airspeed, temperature, heading, altitude, fuel consumption, pressure, destination range, and vertical velocity.
(2) Sample Instructions Information: These are described as semantic chunks, for example, Climb to altitude, Level flight at speed = x, Descend to referenced path, and so on.
(3) Preflight Briefing: Check weather, check flight plan (route, path, etc), check wind speed, and so on.
(4) Taxiing: Checking ground speeds, landmarks, and aircraft queues on the runway.
(5) Take-off: Getting authorization from air traffic control (ATC) tower, checking speed, checking direction, etc.

(6) Cruise Mode: Confirm automation and human roles and types of mode conflicts, envisioning destination on the map display, note position, course, traffic in the air, etc.

Data were also collected from SMEs about task difficulty and suggestions for improvement. The basis of this data collection was to determine what tasks were the hardest to learn and how they could be improved for a pilot in training to learn faster. The SMEs were given a table of tasks and asked to rate each task based on perception of difficulty. The difficulty level was on a Likert scale of 0-5; with 0 being Not difficult and 5 being Most difficult. The tasks were taken from the *Private Pilot Practical Test Standards*, FAA-S-8081-14 (1995), available at http://afts600.faa.gov/data/practicalteststandard/faa-s-8081-14.pdf. The FAA developed the standards for FAA inspectors and designated pilot examiners when conducting pilot practical tests. The particular tasks that the SMEs rated were from Section 1 of the book for the Airplane, Single – Engine Land (ASEL).

After reviewing the SMEs' ratings, some tasks were identified to be most difficult to train, thus, constitute the trainability factors (TF) for VETS design consideration. Trainability factors are tasks that are perceived to score low on the student's achievement metric. The TF as identified by percentage of difficulty rank are:

- Radio communications and ATC (84.3%)
- Normal and Crosswind Approach and Landing (77.6%)
- Short-Field Takeoff and Climb (63.71%)
- Short-Field Approach and Landing (62.5%)
- Go-Around (62.0%)
- Pilotage and Dead Reckoning (60.4%)
- Constant Airspeed Climbs (58.3%)
- Constant Airspeed Descents (56.9%)
- Recovery From Unusual flight Attitudes (56.1%)
- Emergency Approach Landing (54.7%)

These tasks were difficult mostly for cognitive and maneuvering reasons. Other competing tasks for training concerns were maintaining situation awareness in the surrounding airspace, navigating to three-dimensional points in the sky under visual meteorological conditions (VMS), following procedures related to aircraft and airspace operations, and communicating with ATC office and other personnel on the flight deck (Wickens, Gordon, & Liu, 1997). The tasks selected for this study were normal and crosswind approach and landing (NCAL), go-around (GA), constant airspeed climbs, and constant airspeed descends. The last two tasks were combined into one observable task--CSCD. The selected tasks are used by flight instructors to train ab initio flight pilots on desktop computer flight simulators. For example, Hennessy, Wise, and Koonce (1995) selected approach landing tasks to investigate the difference in performance between pathway-in-the-sky display and traditional Instrument Landing System (ILS). Ortiz (1994), in a study of ab initio pilots, used square pattern go-around tasks to compare training effectiveness with and without a PC-based simulator.

# 4. DESCRIPTION OF THE VIRTUAL REALITY SIMULATION TRAINING SOFTWARE

## 4.1 Commercial Off-the-Shelf (COTS) Software Description

Microsoft's "Flight Simulator 2000: Professional Edition" software is used in this experiment. The Professional version of Flight Simulator 2000 is geared to Flight Simulator enthusiasts, real pilots, those who want "more features and more content," and those who are interested in using Flight Simulator 2000 as a PC-based flight training and proficiency aid. The software includes a 3D scenery graphics system. The scenery graphics feature 16-bit color and true elevation data, and is enhanced by textures and seasonal effects. Flight Simulator 2000: Professional Edition, is also optimized for the Intel® Pentium III processor. The software includes various aircraft with instrument panels, virtual cockpits, exterior 3D models and almost all flight data used in every public airport in the world for which an official government agency publishes data. It also displays cities such as London, Paris, New York, Los Angeles, San Francisco, and Chicago in great detail. Flight Simulator 2000 also includes new custom 3D objects, including buildings, vehicles, ships, towers, and more. The Flight Simulator world has incredible realism and immersion.

The weather system provided by the software dramatically improves the variety of weather as a user flies and the effects they see like clouds, precipitation, lightning, and more. A user can customize realistic weather or fly in real-time conditions using the Internet. This is very helpful in adding various complexities to the flight.

For this study, a Software Development Kit (SDK) was used to develop the scenarios for experimentation. The Flight Simulator 2000 Adventure Programming Language (APL) SDK contains documentation and all the necessary components (including a compiler) needed to create any desired scenario. Sound files were recorded and used in conjunction with the three scenarios to provide ATC commands and instructions. The Black Box application runs in conjunction with Flight Simulator 2000. It enables the user to record variables simultaneously such as airspeed, altitude, and heading in 10-second intervals. This unit was developed using Visual C++.

## 4.2 Hardware Requirement

The basic hardware requirement for VETS consists of:

(1) Display: an HMD unit and two monitors. The HMD unit and one of the monitors are used to display the same image frames at all times enabling the instructor and other students to observe the process. The other monitor displays the main operating system and control functions, which will not be shown on the HMD unit. This setup permits the instructor to have full control of the VR environment such as changing tasks, flight parameters, and so on.
(2) Input device: keyboard, mouse, and joystick
(3) Output device: printer and speaker system

Figure 2 graphically portrays this configuration.

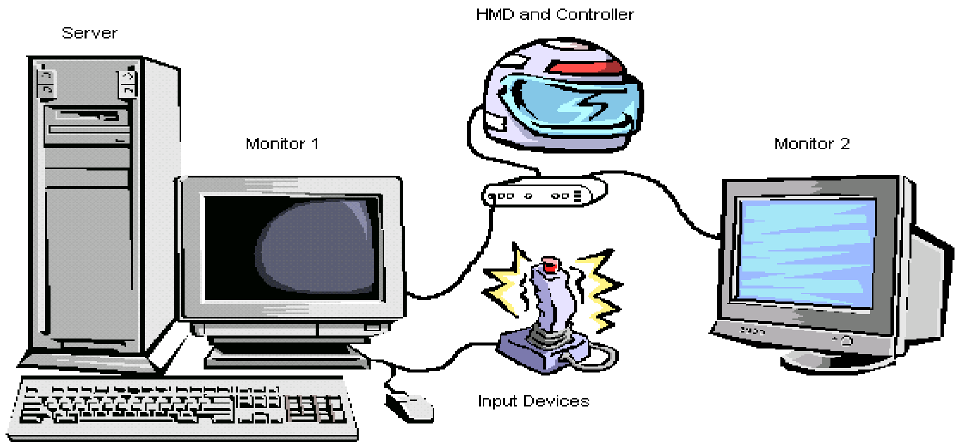


Figure 2. Hardware Configuration for VETS

## 4.3 Software

The VETS software was developed from a cluster of Engineering Animation World Up, Version 4; and Spatial Technology's 3D Studio Max, Release 3.1. Both software allow for design and modification of any aircraft geometry and aerodynamic characteristics. The simulation graphics can be displayed on a computer monitor and/or HMD. The HMD refresh rate is set at 60 Hz for the current experiment. Displaying graphics both on the monitor(s) and HMD enables instructors to give verbal commands and to monitor the student pilot's performance on site.

# 5. EXPERIMENTAL DESIGN

## 5. 1 Preamble

This experiment investigated training of piloting skills in immersive and nonimmersive virtual environments. Participants were asked to complete three piloting tasks in five trials each. The overall question of the experiment is: Do IVEs enhance learning and task performance more than INVEs for any of the experimental tasks?

## 5.2 Method

### 5.2.1 Participants

Thirty subjects participated in the experiment. They consisted of undergraduate and graduate students from North Carolina A&T State University and the general public. The age range was between 18 and 50 years. The participants were randomly assigned to two counterbalanced groups: IVE and NIVE, respectively.

### 5.2.2 Measures

Independent variables: The independent measures manipulated were the environment (IVE and NIVE), tasks, and number of training trials.

Dependent variables: The major measures of interest for this experiment were errors and error rates for altitude, airspeed, vertical speed, and heading. The standards for measurement comparison are acceptable ranges of flight set by the flight simulator software as recommended by flight instructors.

### 5.2.3 Apparatus

The NIVE apparatus comprised of one personal computer with the following hardware specifications: one Intel 733 MHz Pentium Processor, one video card with 1228 MB of RAM, one 17-inch SVGA monitor, one 30GB hard-disk drive and one joystick. It also includes one copy of Microsoft's Flight Simulator 2000: Professional Edition software, and one set of 20WPC loud speakers (Figure 3).

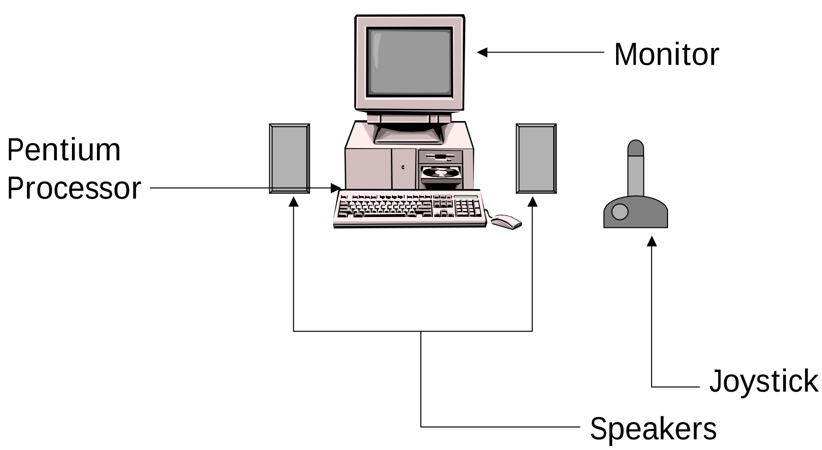


Figure 3. Nonimmersive Virtual Environment

The IVE apparatus comprised of the same features as the NIVE but included a few more features. They were one set of 32-ohm headphones, Pro-Logic Sound Amplifier, and a HMD (Figure 4). The HMD was compatible for eyeglasses and had 100% overlap. The field-of-view was 35$^{o}$ diagonal with total field-of-regard of 21$^{o}$ (V) X 28$^{o}$ (H). The HMD had full XGA resolution of 1024 horizontal pixels by 768 vertical lines. Participants in this experiment did not wear a head tracker.

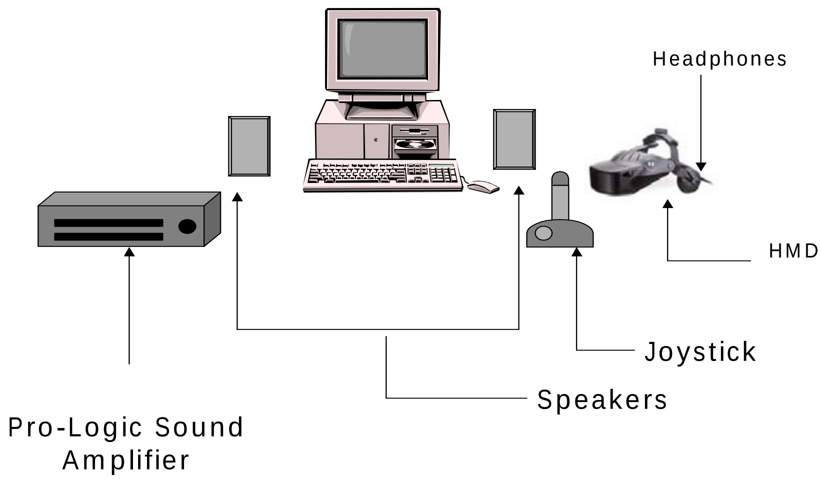


Figure 4: Immersive Virtual Environment

The virtual environment software emulated the cockpit of a Cessna airplane. It simulated contact flight, instrument flight, various terrains, wind, clouds, and turbulence, to name a few (see Figure 5). The subjects were asked to focus on the airspeed indicator, magnetic compass, altimeter, vertical airspeed, flap positioner, pitch and trim.


Figure 5. Sample Cockpit Display

The user interface was similar for both displays. In the NIVE the subjects used the joystick and, in the IVE the participants used the HMD and the joystick. The joystick is used to navigate through the VE. The joystick simulated the flight controls for pitch, roll, yaw, and power.

**5.2.4 Procedure.** We divided the experiment into five sections:

(1) Introduction: In the introduction phase, the experimenter explained the purpose of the experiment and the risk involved.

(2) Training: In this phase, the subjects were introduced to the cockpit layout, instruments, and displays. Only the relevant cockpit instruments needed for flying tasks were elaborated. The subjects also learned to use the joystick for navigation.

(3) Preliminary Learning of Flight Task Scenarios: In this phase, the subjects were introduced to the three flying tasks: NCAL--this task tested the subject's ability to approach and land on a runway with a moderate wind; GA was the second task--it involved the subjects flying the plane around the runway as if it were a missed approach in order to land; and CSCD, in which the subject performed sample climbing or descending at a given speed. Figure 6 shows sample screen capture of the trial tasks.

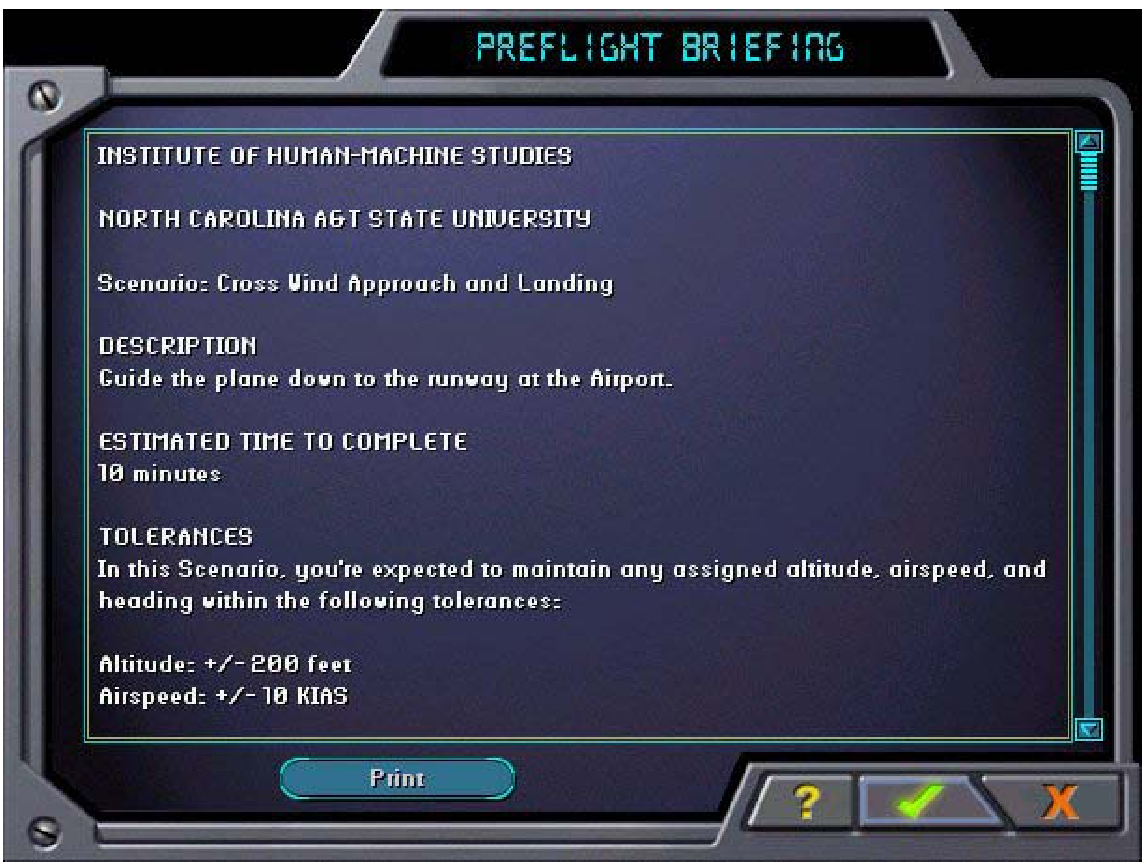


Figure 6. Preflight Briefing for Normal Crosswind Approach and Landing

(4) Actual Flight Experiments: Each subject was supervised to perform a total of 15 tests (in 3 blocks of 5 tests each); the three blocks were separated into three days of tests with one block randomly assigned each day. The experiment took an average of 2.83 hours (with standard deviation of 0.74 hours) to complete per participant.

The three experimental blocks consist of the three flight tasks: NCAL, GA, and CSCD. The simulated ATC voice commands were randomly called to the subjects. At the end of each test, the subjects were asked to provide an after-fact debrief by filling out a scenario experience questionnaire (SEQ) After the first test, the experimenter proceeded with other tests without any more questioning. The procedure was repeated for subsequent test scenarios.

### 5.2.5 Data Collection

The main data on the subject profiles and flight performance measurements were automatically collected by the VETS software during experimental trial. The data collection module was developed with Microsoft Excel™. Pertinent information for each subject included:

- Name
- Age
- Gender (Male (M) or Female (F))
- Flight Simulator Experience (1 – Yes, 2 - No)
- Piloting Experience (1 - Yes, 2 - No)
- Virtual Environment (1 - NIVE, 2 - IVE)
- Task (1 - Crosswind Approach and Landing; 2 - Go-Around; 3 - Constant Airspeed with Changes in Altitude during Climbing or Descending)
- Flight Variables Measured: – ALT – altitude (feet)
  – HDG – heading (degrees)
  – AS – airspeed (knots)
  – VAS – vertical airspeed (100ft/min)
- ER – Error Rate (#errors/# of 10 second intervals)
- ERR –Number of errors observed

Altitude, heading, airspeed, and vertical speed were recorded at 10-second intervals during each flight scenario for each participant. A simple computer program was written to calculate the total number of errors and error rate for each observation during a trial scenario. The error difference was computed by subtracting observed performance data by participants from perfect scenario data (standard) recommended by flight instructors and flown by the Flight Simulator autopilot. From the error difference values, the mean altitude, heading, airspeed, and position error rates for each variable were calculated by. The perfect or standard scenario data recommended by expert flight instructors (EFI) were selected from the FAA Practical Test Standard FAA-S-8081-14 handbook. The flight performance limits are:

- Altitude (+/- 200 ft)
- Heading (+/- 20 degrees)
- Airspeed (+/-10 knots)
- VAS (+/- 1000 ft/min)

### 5.2.6 Sample Data Collection

Table 1 shows sample data on averages and standard deviations for altitude errors for both NIVE groups (subjects 1-15) and IVE groups (subjects 16-30).

Table 1. Sample Altitude Error Averages and Standard Deviations for NCAL.

| NIVE | | |
|---|---|---|
| SUBJECT | AVG | STDEV |
| 1 | 0.2662 | 0.2467 |
| 2 | 0.5201 | 0.2400 |
| 3 | 0.5514 | 0.1747 |
| 4 | 0.3649 | 0.2882 |
| 5 | 0.4493 | 0.1523 |
| 6 | 0.5083 | 0.2306 |
| 7 | 0.3832 | 0.0755 |
| 8 | 0.4704 | 0.2460 |
| 9 | 0.4756 | 0.0364 |
| 10 | 0.3478 | 0.2420 |
| 11 | 0.4926 | 0.1247 |
| 12 | 0.5143 | 0.0453 |
| 13 | 0.5205 | 0.1860 |
| 14 | 0.2443 | 0.3125 |
| 15 | 0.5394 | 0.2272 |
| IVE | | |
| SUBJECT | AVG | STDEV |
| 16 | 0.5035 | 0.0993 |
| 17 | 0.5446 | 0.0439 |
| 18 | 0.4975 | 0.1028 |
| 19 | 0.5192 | 0.2152 |
| 20 | 0.5806 | 0.1325 |
| 21 | 0.4563 | 0.2888 |
| 22 | 0.3227 | 0.2487 |
| 23 | 0.5413 | 0.0464 |
| 24 | 0.4691 | 0.1694 |
| 25 | 0.5731 | 0.1523 |
| 26 | 0.4419 | 0.1545 |
| 27 | 0.5654 | 0.0709 |
| 28 | 0.3635 | 0.1802 |
| 29 | 0.3900 | 0.1698 |
| 30 | 0.5580 | 0.1003 |

### 5.2.7 Determining Error Rates

The error rates were calculated by dividing the number of samples exceeding performance criteria for each variable by the total number of samples recorded at 10-second intervals over the trial. A maximum time of 10 inutes was given to participants for completion of each task scenario, yielding a possible maximum 60 data points per task trial. The number of intervals for each participant for any given task was different based on individual performance times

The number of errors were then counted and recorded for each variable and divided by the total number of 10-second intervals performed by the user for the task to obtain the error rate. An example illustration is shown in Figure 7. The colored columns gives sample error values.

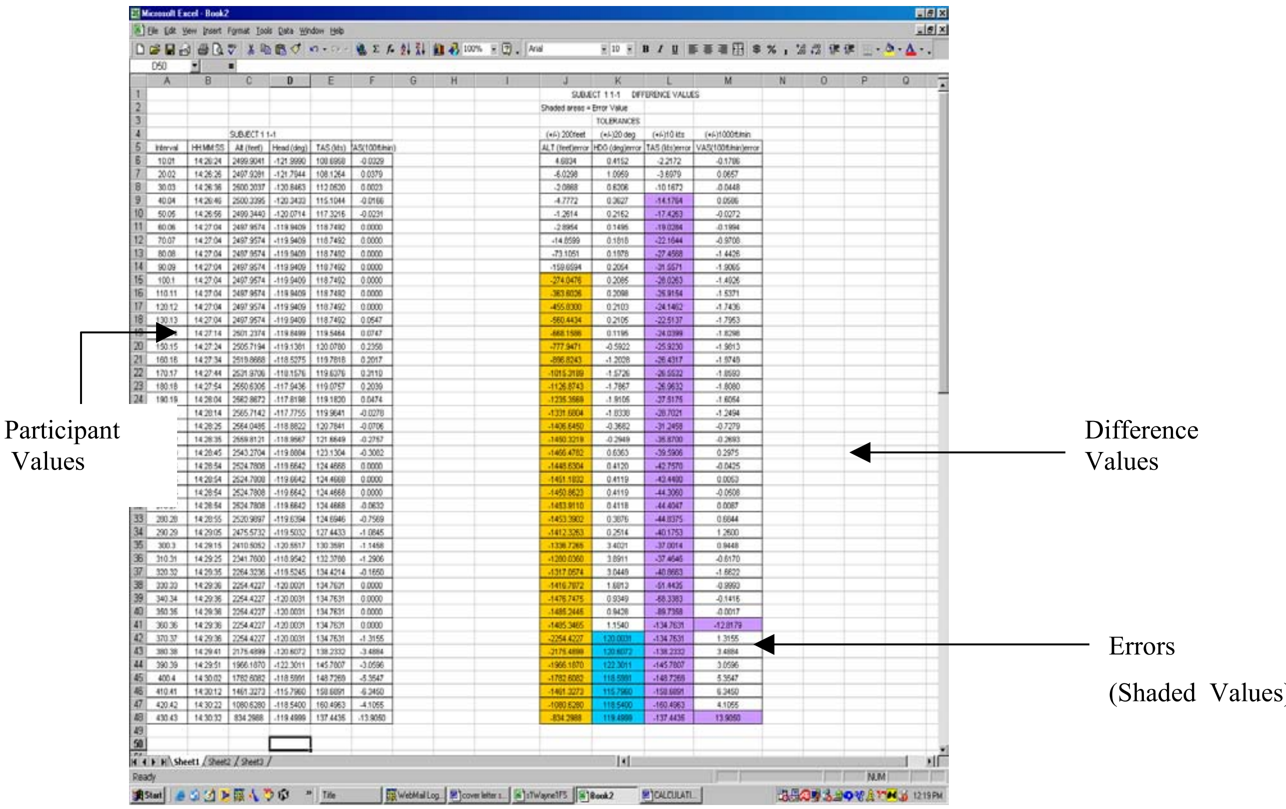


Figure 7. Sample Error and Error Differences used in Error Rate Calculation

Table 2 gives an example error difference derived for NCAL task under NIVE for one subject during a single trial.

Table 2. Sample Error Difference NCAL Task using NIVE.

SUBJECT 1 1-1 DIFFERENCE VALUES

Shaded areas are samples exceeding the error tolerance

TOLERANCES

| (+/-) 200 feet | (+/-)20 deg | (+/-)10 kts | (+/-)1000 ft/min |
|---|---|---|---|
| ALT (feet)error | HDG (deg)error | TAS (kts)error | VAS(100ft/min)error |
| 4.6834 | 0.4152 | -2.2172 | -0.1786 |
| -6.0298 | 1.0959 | -3.6979 | 0.0657 |
| -2.0868 | 0.6206 | -10.1672 | -0.0448 |
| -4.7772 | 0.3627 | -14.1764 | 0.0586 |
| -1.2614 | 0.2162 | -17.4263 | -0.0272 |
| -2.8954 | 0.1495 | -19.0284 | -0.1994 |
| -14.8599 | 0.1818 | -22.1644 | -0.9708 |
| -73.1051 | 0.1978 | -27.4568 | -1.4426 |
| -159.6594 | 0.2054 | -31.5571 | -1.9065 |
| -274.0476 | 0.2085 | -28.0263 | -1.4926 |
| -363.6026 | 0.2098 | -25.9154 | -1.5371 |
| -455.8300 | 0.2103 | -24.1462 | -1.7436 |
| -560.4434 | 0.2105 | -22.5137 | -1.7953 |
| -668.1586 | 0.1195 | -24.0399 | -1.8298 |
| -777.9471 | -0.5922 | -25.9230 | -1.9813 |
| -896.8243 | -1.2028 | -26.4317 | -1.9749 |
| -1015.3189 | -1.5726 | -26.5532 | -1.8593 |
| -1126.8743 | -1.7867 | -26.9632 | -1.8080 |
| -1235.3569 | -1.9105 | -27.5175 | -1.6054 |
| -1331.6804 | -1.8338 | -28.7021 | -1.2494 |
| -1406.6450 | -0.3682 | -31.2458 | -0.7279 |
| -1450.3219 | -0.2949 | -35.8700 | -0.2693 |
| -1466.4782 | 0.6363 | -39.5906 | 0.2975 |
| -1448.6304 | 0.4120 | -42.7570 | -0.0425 |
| -1451.1832 | 0.4119 | -43.4480 | 0.0053 |
| -1450.8623 | 0.4119 | -44.3060 | -0.0508 |
| -1453.9110 | 0.4118 | -44.4047 | 0.0087 |
| -1453.3902 | 0.3876 | -44.8375 | 0.6844 |
| -1412.3263 | 0.2514 | -40.1753 | 1.2600 |
| -1336.7265 | 3.4021 | -37.0014 | 0.9448 |
| -1280.0360 | 3.8911 | -37.4646 | -0.6170 |
| -1317.0574 | 3.0449 | -40.8663 | -1.6622 |
| -1416.7872 | 1.6813 | -51.4435 | -0.9993 |
| -1476.7475 | 0.9349 | -68.3383 | -0.1416 |
| -1485.2445 | 0.9428 | -89.7358 | -0.0017 |
| -1485.3465 | 1.1540 | -134.7631 | -12.8179 |
| -2254.4227 | 120.0031 | -134.7631 | 1.3155 |
| -2175.4899 | 120.6072 | -138.2332 | 3.4884 |
| -1966.1870 | 122.3011 | -145.7807 | 3.0596 |
| -1782.6082 | 118.5991 | -148.7269 | 5.3547 |
| -1461.3273 | 115.7960 | -158.6891 | 6.3450 |

# 6. DATA ANALYSIS

## 6.1 Test Question 1: Are there statistically significant differences in task performance error rates between subjects trained under IVE and NIVE?

A Two-Way Mixed Analysis of Variance (ANOVA) technique was used to analyze the data for error rates.. There was one between-subjects factor, which was the type of training (NIVE and IVE). There was also one within-subjects factor, which was trial (the five trials within each task) completed by each participant. The data were analyzed with the SAS software package. The differences among tasks and between measures within a task were not compared.

### 6.1.1 Results for Error Rate Comparisons

Table 3 and 4 show the significant effects for each task for error rates and error with a level of significance of $\alpha = 0.5$. The "X" in Table 5 indicates significance.

Table 3: Significant Effects Table for Error Rates.

| | | ERROR RATE TYPE | | | |
|---|---|---|---|---|---|
| **TASK** | **EFFECT** | **ALTITUDE** | **HEADING** | **AIRSPEED** | **V.AIRSPEED** |
| | | | | | |
| 1 (NCAL) | Environment | | | | |
| | Training | X | X | | |
| | Interaction | | | | |
| | | | | | |
| 2 (GA) | Environment | X | X | | |
| | Training | | | | |
| | Interaction | X | X | | |
| | | | | | |
| 3 CSCD) | Environment | | | | |
| | Training | | | | |
| | Interaction | | | | |

6.1.1.1 Altitude Error Rates

Performance for controlling the correct altitude on the NCAL task was affected by training trials ($F(4,28) = 2.49$, $p < 0.013$); GA task was affected by the training environment ($F(1,4) = 4.59$, $p < 0.0411$). There was also interaction between environment and training trials ($F(4,28) = 7.12$, $p < 0.0001$) as shown in Figure 8. The interaction revealed that NIVE showed an immediate and sustained reduction in error rates after the first training trial; on the other hand, IVE subjects seem to make an increasing error rate after the first trial. CSCD task error rate was not affected by either training trials or task environment (IVE or NIVE). In general, NIVE subjects performed better with respect to error rate on GA altitude tasks. There were no statistically significance differences in NCAL and CSCD altitude error rates.

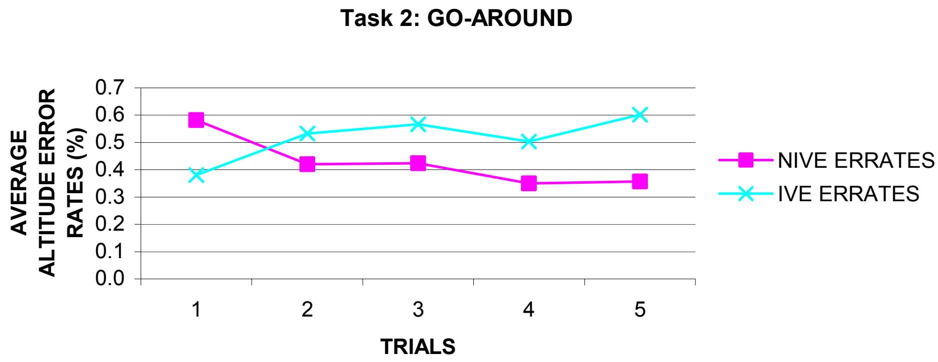


Figure 8. Interaction between Training Trials and Training Environment in Altitude Control Error Rate for the GA Task

6.1.1.2 Heading Error Rates

Performance for controlling heading on the NCAL task was affected by training trials ($F(4,28) = 2.93$, $p < 0.05$). Performance for controlling heading on the GA task was affected by the environment ($F(1,4) = 4.59$, $p < 0.0411$). There was also interaction between environment and training ($F(4,112) = 8.78$, $p < 0.025$) as shown in Figure 9. NIVE showed decreasing and sustained error rates at and after the second trial; IVE subjects maintain increasing error rate after the first trial. Overall, NIVE subjects performed better than the IVE group in GA heading error rates. There were no statistically significance differences in NCAL and CSCD altitude error rates.

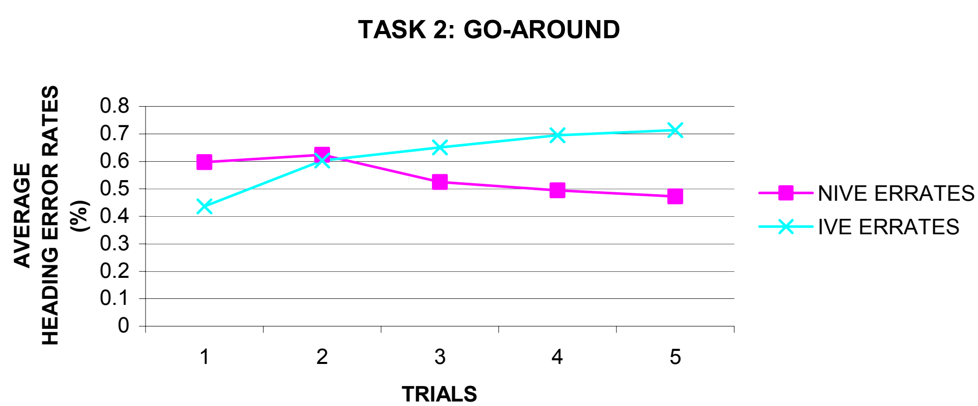


Figure 9. Interaction between Training Trials and Training Environment in Heading Control Error Rate for the GA task

6.1.1.3 Airspeed and Vertical Airspeed Error Rates

There was no statistical significance difference between NIVE and IVE in error rate performance for all three tasks (experimental scenarios).

**6.2 Test Question 2: Are there statistically significant differences in task performance error between subjects trained under IVE and NIVE?**

### 6.2.1 Results from Error Comparisons

Table 4: Significant Effects Table for Error.

| | | ERROR TYPE | | | |
|---|---|---|---|---|---|
| TASK | EFFECT | ALTITUDE | HEADING | AIRSPEED | V.AIRSPEED |
| | | | | | |
| 1 (NCAL) | Environment | | X | X | |
| | Training | X | X | X | |
| | Interaction | X | X | | |
| | | | | | |
| 2 (GA) | Environment | | | | |
| | Training | | | | |
| | Interaction | X | X | X | |
| | | | | | |
| 3 (CSCD) | Environment | X | X | | |
| | Training | X | | | |
| | Interaction | | | | |

6.2.1.1 Altitude Error

Performance for altitude control on the NCAL task was affected by training trials ($F (4, 29) = 2.77, p < 0.0306$). There was also an interaction between environment and training trials ($F (4,111) = 3.73, p < 0.0069$) as shown in Figure 10. The interaction reveals that, overall, NIVE subjects performed with higher error on the first trial than the IVE group; but the groups performed comparably on trials 2–5.

GA tasks also showed interaction effect between the environment and training ($F(4,112) = 5.39, p < 0.005$) as shown in Figure 11. Figure 11 reveals that the NIVE subjects had greater errors on the first trial but performance was not different from IVE subjects afterwards. CSCD task was affected by the environment ($F(1,4) = 6.08, p < 0.0200$) and training ($F (4,28) = 2.53, p < 0.0444$), but no interaction was observed.

Overall, if the first trial effect is excluded, the NIVE group performed better with NCAL and GA altitude tasks performance with respect to average number of errors. No statistical significance difference was observed for CSCD altitude tasks.

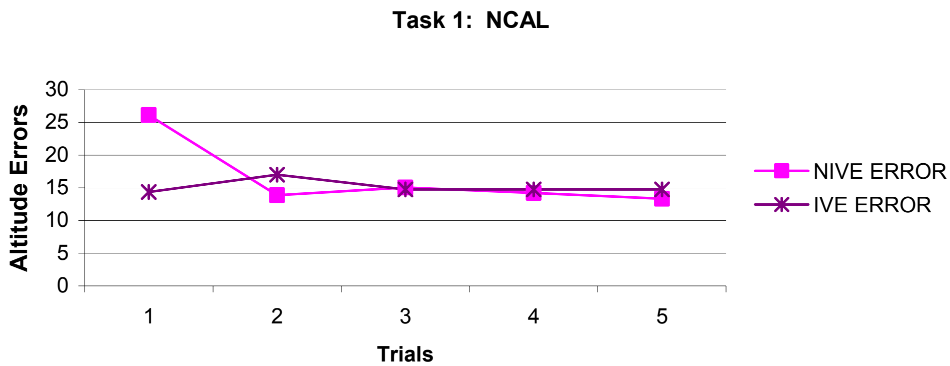


Figure 10: Interaction between Trials and Training Environment for Altitude Control Error on the NCAL Task

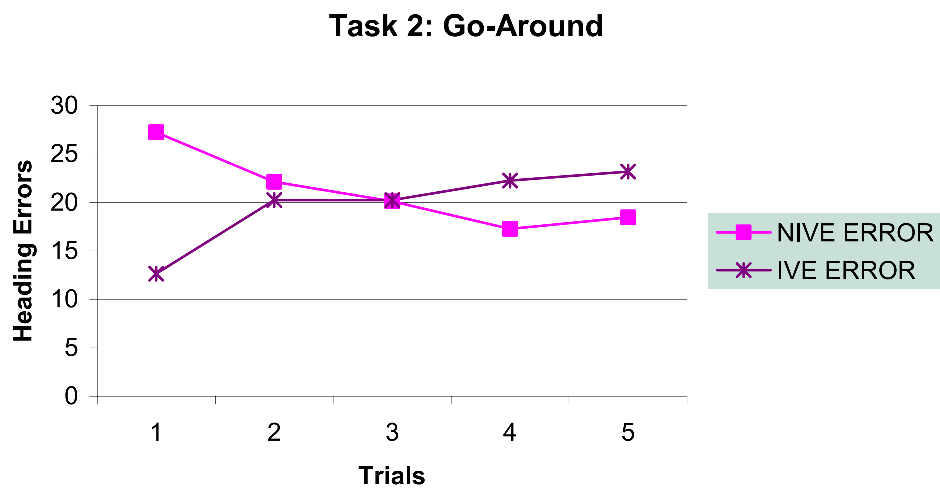


Figure 11: Interaction between Trials and Training Environment in Heading Control Error for the GA Task

6.2.1.2 Heading Errors

Performance for heading control on the NCAL task was affected by the environment (F (1,4) = 4.74, p < 0.0381) and training (F (4,28) = 3.11, p < 0.0182). There was also interaction between environment and training (F (4,112) = p < 0.0160) as shown in Figure 12.

The NIVE group experienced stable reduction on errors during and after the 2nd trial; the IVE group experienced early high performance (reduced error) but this performance degraded after the 3rd trial. This is an interesting trend: IVE group performed better during the 1st, 2nd, and 3rd trials, and degraded after; the NIVE group degraded earlier during 1st, 2nd, and 3rd trials, and degraded thereafter.

The GA task showed interaction between environment and training (F (4,112) = 5.57, p < 0.0004) as shown in Figure 13. The NIVE group tended to exhibit higher errors on the first training trial but similar performance to the IVE group for following trials. CSCD task was affected by the environment (F 1,4) = 6.25, p<0.0185), but no interaction between the environment and training was observed.

Overall, the IVE group performed better than NIVE group in NCAL heading tasks, and NIVE group performed better in GA heading tasks. There was no significance difference in error between NIVE and IVE group in CSCD heading tasks.

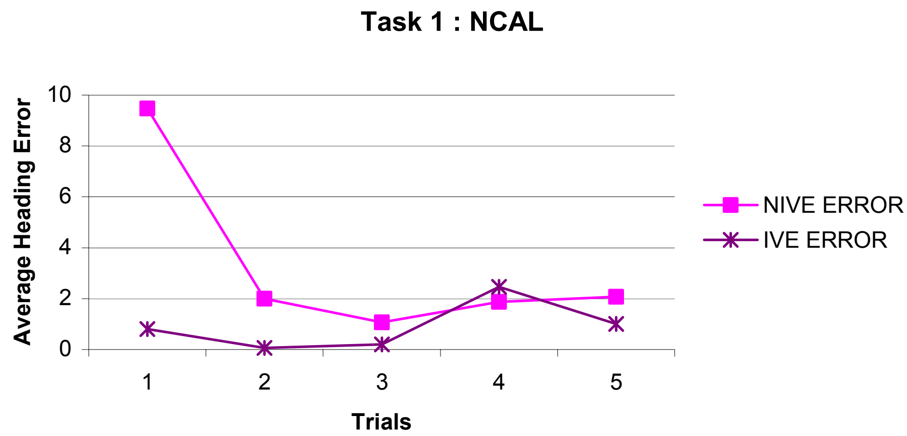


Figure 12 : Interaction between Trials and Training Environment for Heading Control Error for the NCAL task

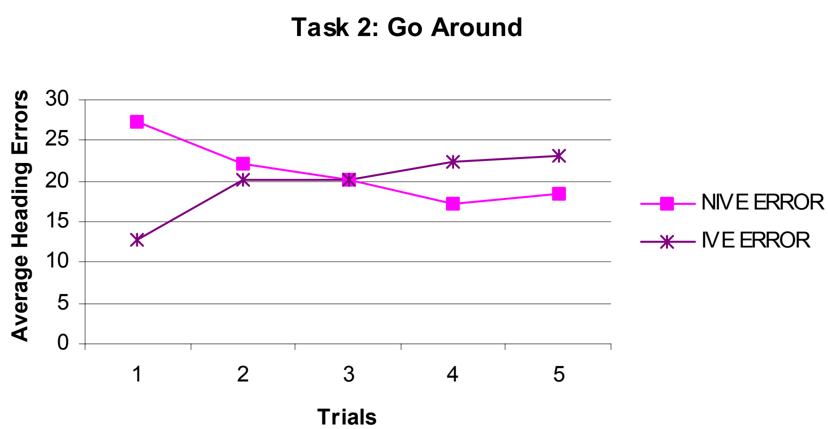


Figure 13: Interaction between Trials and Training Environment for Heading Control Error Rate on the GA task

6.2.1.3. Airspeed Error

Performance for airspeed control on the NCAL task was affected by the environment (F (1,4) = 16.03, p < 0.0004) and training (F (4,28) = 3.72, p < 0.0070), but there was no significant interaction between environment and training trials. Performance for airspeed control on the NCAL task was affected by the environment, $\overline{X}_{IVE} = 123$, $\overline{X}_{NIVE} = 456$, (F (1,4) = 16.03,

$p < 0.0004$). This indicated that IVE group performed better in terms of minimum average error than the NIVE group. There was no statistical significant difference between NIVE and IVE groups in terms of error distribution for the five trials.

Performance for airspeed control for the GA task showed an interaction effect between environment and training trials ($F(4,112) = 2.72$, $p < 0.0333$) as shown in Figure 14 ($\overline{X}_{IVE} = 22.43 = \overline{X}_{NIVE}$ at 2nd trial, and ($\overline{X}_{IVE} = 19.32 = \overline{X}_{NIVE}$ at 4th trial. The NIVE group showed a level-off improvement during and after the first trial.

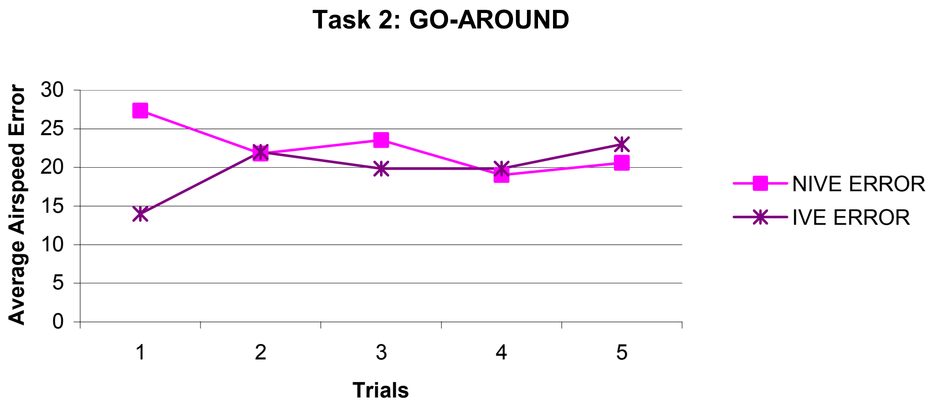


Figure 14: Interaction Between Trials and Training Environment for Airspeed Control Error for the GA task

6.2.1.4 Vertical Airspeed Error. There was no significance effect observed either by the environment or training in any of the tasks.

### 6.3 Test Question 3: Does IVE provide better pilot skill training than NIVE?

In this test, we compare performance improvement of error rates across task trials, between the first trial and the last trial for each task, and finally, between the last trials for both IVE and NIVE. The student t-tests was used for the analysis. The results were obtained at a level of significance of $\alpha = 0.5$ ( $t_{.975}(14) = 2.673$).

#### 6.3.1 Error Rate Analysis beginning (1st trial) and end (5th trial) for NCAL Tasks

As shown in Table 5, under NIVE, there was a significant difference between the first and second trial performance for altitude error rate ($t(14) = 2.82 > 2.673$; $p = 0.019$). For calculated p values << 0.001, there were no significance differences between the 1st and 5th trials for heading, airspeed, and vertical airspeed errors under NIVE. With IVE group , there were no significant differences in all measured error rates. These results indicate no performance improvement in the 5th trial for all the performance variables. The NIVE group performed better than the IVE group for altitude error rate.

Table 5: The Summary of Statistics for 1st and 5th Trials for NCAL Tasks.

| | | Training Environment | | | |
|---|---|---|---|---|---|
| | | NIVE | | IVE | |
| | Sample | 15 | | 15 | |
| Altitude | Begin | 0.5998 | 0.249** | 0.484 | 0.211** |
| | End | 0.3606 | 0.215 | 0.4307 | 0.194 |
| | Average | 0.444 | | 0.488 | |
| | t-statistics | **2.822** | | 0.409 | |
| Heading | Begin | 0.186 | 0.196 | 0.052 | 0.138 |
| | End | 0.044 | 0.147 | 0.09 | 0.008 |
| | Average | 0.078 | | 0.03 | |
| | t-statistics | 0.142 | | -1.106 | |
| Airspeed | Begin | 0.546 | 0.192 | 0.518 | 0.163 |
| | End | 0.524 | 0.119 | 0.422 | 0.219 |
| | Average | 0.526 | | 0.468 | |
| | t-statistics | 0.377 | | 1.367 | |
| V. Airspeed | Begin | 0.02 | 0.013 | 0.024 | 0.019 |
| | End | 0.028 | 0.02 | 0.038 | 0.015 |
| | Average | 0.029 | 0.03 | 0.468 | |
| | t-statistics | -0.325 | | -0.01 | |

Note: ** = values of standard deviation as second pair

Further analysis was performed to compare the 5th trial error rates for both NIVE and IVE The results of the analysis is shown in Table 6. There were no significance differences in error rate performance for altitude and heading tasks. For airspeed error rate, IVE group performed better in the 5th trial than the NIVE group ($t = 2.979$, $p < 0.017$; $\overline{X}_{IVE} = 0.422$, $\overline{X}_{NIVE} = 0.524$). On the other hand, the NIVE group showed marginal improvement over IVE group in vertical airspeed error rate ( $t = 1.727$; $p < 0.027$; $\overline{X}_{IVE} = 0.038$, $\overline{X}_{NIVE} = 0.028$).

Figure 15 is used to show the average error rate performance for NCAL tasks per the discussions.

As shown in Figure 15, overall, there were no statistical differences in error rate performance between NIVE and IVE groups for vertical airspeed and altitude error rates. Statistical differences were observed for heading error rate ($t = 2.781$, p, 0.001; $\overline{X}_{IVE} = 0.03$ $\overline{X}_{NIVE} = 0.08$), and altitude error rate ($t = 3.09$; $p < 0.00001$; $\overline{X}_{IVE} = 0.49$, $\overline{X}_{NIVE} = 0.44$).

Table 6: Comparing 5th Trial Error Performance for NCAL Tasks

| | | Training Environment NIVE | IVE | |
|---|---|---|---|---|
| | Sample | 15 | 15 | |
| Altitude | Average | 3.606 | 0.4307 | |
| | std. | 0.215 | 0.194 | t = -1.228 |
| Heading | Sample | 15 | 15 | |
| | Average | 0.044 | 0.09 | |
| | std. | 0.147 | 0.03 | t = -1.210 |
| Airspeed | Sample | 15 | 15 | |
| | Average | 0.524 | 0.422 | |
| | std. | 0.19 | 0.219 | **t = 2.979** |
| V. Airspeed | Sample | 15 | 15 | |
| | Average | 0.028 | 0.038 | **t = 1.727** |

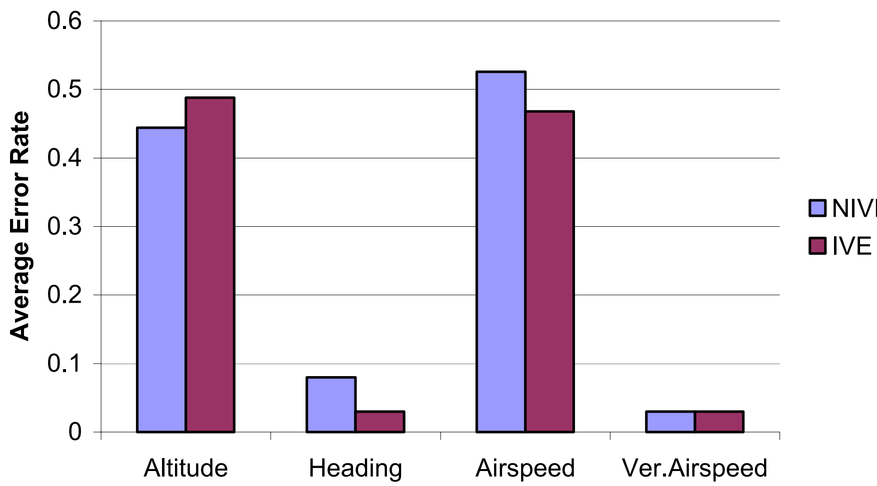


Figure 15: Average Error Rate Distribution for NCAL Tasks.

### 6.3.2 Error Rate Analysis beginning (1st trial) and end (5th trial) for GA Tasks

As shown in Table 7, under NIVE, there was a significant difference between 1st and 2nd trial performance for altitude error rate ($t = 2.78$; $p = 0.014$). For calculated p values << 0.001, there were no significance differences between 1st and 5th trials for heading, airspeed, and vertical airspeed errors with the NIVE group.

With the IVE group, there were significant differences in altitude error rate ($t = -3.367$, $p = 0.043$), heading error rate ($t= -5.72$, $p = 0.0002$), airspeed ( $t = -2.907$, $p = 0.126$). These results indicate no performance improvement in the 5th trial using IVE for this performance variables.

However, in the IVE group, there was a noticeable statistical difference in improvement between the 1st and 5th trials for vertical airspeed error rate ($t = 2.517$, $p < 0.0001$). Here, the IVE group performed better than the NIVE group who showed no performance gain.

Table 7: The Summary of Statistics for 1st and 5th Trials for GA Tasks.

| | | Training | Environment | | |
|---|---|---|---|---|---|
| | | NIVE | | IVE | |
| | Sample | 15 | | 15 | |
| Altitude | Begin | 0.582 | 0.244** | 0.38 | 0.216** |
| | End | 0.357 | 0.199 | 0.602 | 0.09 |
| | Average | 0.427 | | 0.517 | |
| | t-statistics | **2.78** | | **-3.674** | |
| Heading | Begin | 0.597 | 0.305 | 0.436 | 0.145 |
| | End | 0.472 | 0.212 | 0.714 | 0.12 |
| | Average | 0.542 | | 0.619 | |
| | t-statistics | 1.642 | | **-5.72** | |
| Airspeed | Begin | 0.623 | 0.244 | 0.52 | 0.218 |
| | End | 0.534 | 0.159 | 0.709 | 0.126 |
| | Average | 0.581 | | 0.621 | |
| | t-statistics | 1.184 | | **-2.907** | |
| V. Airspeed | Begin | 0.021 | 0.035 | 0.026 | 0.04 |
| | End | 0.035 | 0.07 | 0 | 0 |
| | Average | 0.033 | | 0.009 | |
| | t-statistics | -0.594 | | **2.517** | |

Note: ** = values of standard deviation as second pair

Further analysis was performed to compare the 5th trial error rates for both NIVE and IVE. The result of the analysis is shown in Table 8. There were significance differences in error rate performance for all tasks. The NIVE group performed better for altitude error rate ($t = -4.74$, $p < 0.001$; $\overline{X}_{IVE} = 0.709$, $\overline{X}_{NIVE} = 0.72$) and airspeed error rate ($t = -4.17$, $p < 0.0032$; $\overline{X}_{IVE} = 0.709$) while the IVE group performed better for heading error rate ($t = 4.36$, $p < 0.0001$; $\overline{X}_{IVE} = 0.714$, $\overline{X}_{NIVE} = 0.72$) and vertical airspeed ($t = 1.92$, $p < 0.033$; $\overline{X}_{IVE} = 0.0$, $\overline{X}_{NIVE} = 0.035$).

Figure 16 is used to show the average error rate performance for GA tasks per the discussions. As shown in Figure 16, there were no statistical significance differences in error rate performance between NIVE and IVE groups for heading and airspeed error rates. Statistical differences were observed for altitude error rate ($t = -2.833$, p, 0.042; $\overline{X}_{IVE} = 0.517$ $\overline{X}_{NIVE} = 0.427$), and vertical airspeed error rate ($t = -2.92$; $p < 0.00001$; $\overline{X}_{IVE} = 0.009$, $\overline{X}_{NIVE} = 0.033$). The results indicate that NIVE may be good for training GA-related altitude tasks while IVE may be appropriate for vertical airspeed-related tasks.

Table 8: Comparing 5th Trial Error Performance for GA Tasks

| | | Training Environment NIVE | IVE | |
|---|---|---|---|---|
| | Sample | 15 | 15 | |
| Altitude | Average | 0.357 | 0.602 | |
| | std. | 0.199 | 0.09 | **t=-4.74** |
| Heading | Sample | 15 | 15 | |
| | Average | 0.72 | 0.714 | |
| | std. | 0.212 | 0.12 | **t = -4.36** |
| Airspeed | Sample | 15 | 15 | |
| | Average | 0.534 | 0.709 | |
| | std. | 0.159 | 0.126 | **t = -4.17** |
| V. Airspeed | Sample | 15 | 15 | |
| | Average | 0.035 | 0 | |
| | | 0.007 | 0.009 | **t = 1.92** |

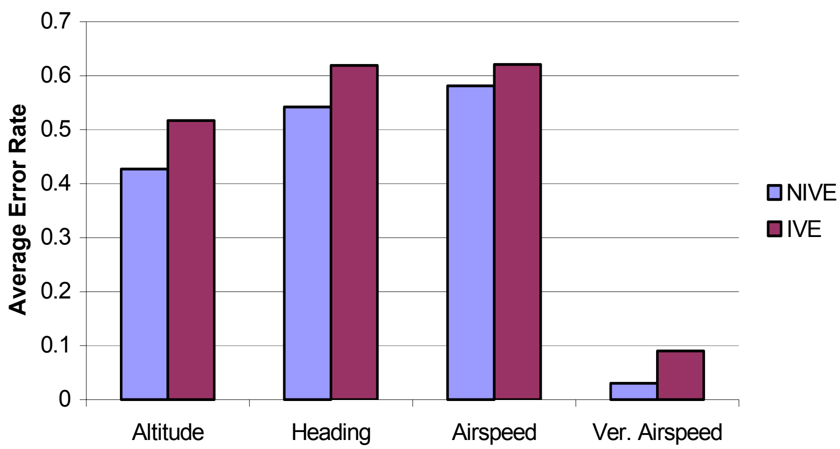


Figure 16: Average Error Rate Distribution for GA Tasks

### 6.3.3 Error Rate Analysis beginning (1st trial) and end (5th trial) for CSCD Tasks

As shown in Table 9, There were no noticeable statistical differences between NIVE and IVE groups in performance gains. That is, we accept the null hypotheses that for CSCD tasks, there is no change in performance using either NIVE or IVE training.

Further analysis was performed to compare the 5th trial error rates for both NIVE and IVE. The result of the analysis is shown in Table 10. Again, there were no noticeable performance differences between the IVE and NIVE group. Figure 17 illustrates these results.

Table 9 :The Summary of Statistics for 1st and 5th Trials for CSCD Tasks.

| | | Training | Environment | | |
|---|---|---|---|---|---|
| | | NIVE | | IVE | |
| | Sample | 15 | | 15 | |
| Altitude | Begin | 0.524 | 0.16** | 0.456 | 0.219** |
| | End | 0.448 | 0.13 | 0.345 | 0.197 |
| | Average | 0.463 | | 0.383 | |
| | t-statistics | 1.428 | | 1.443 | |
| Heading | Begin | 0.328 | 0.187 | 0.317 | 0.2 |
| | End | 0.298 | 0.2 | 0.255 | 0.23 |
| | Average | 0.304 | | 0.233 | |
| | t-statistics | 0.424 | | 0.788 | |
| Airspeed | Begin | 0.588 | 0.211 | 0.448 | 0.187 |
| | End | 0.472 | 0.27 | 0.421 | 0.187 |
| | Average | 0.512 | | 0.48 | |
| | t-statistics | 1.311 | | 0.395 | |
| V. Airspeed | Begin | 0.034 | 0.013 | 0.144 | 0.26 |
| | End | 0.046 | 0.06 | 0.044 | 0.06 |
| | Average | 0.058 | | 0.066 | |
| | t-statistics | -0.756 | | 1.451 | |

Note: ** = values of standard deviation as second pair

Table 10: Comparing 5th Trial Error Performance for CSCD Tasks

| | Training | Environment | |
|---|---|---|---|
| | NIVE | IVE | |
| Sample | 15 | 15 | |
| Average | 0.448 | 0.345 | |
| std. | 0.13 | 0.197 | t = 1.69 |
| | | | |
| Sample | 15 | 15 | |
| Average | 0.298 | 0.255 | |
| std. | 0.12 | 0.23 | t = 0.546 |
| | | | |
| Sample | 15 | 15 | |
| Average | 0.472 | 0.421 | |
| std. | 0.27 | 0.187 | t = 0.601 |
| | | | |
| Sample | 15 | 15 | |
| Average | 0.046 | 0.044 | |
| | 0.06 | 0.06 | t = 0.09 |

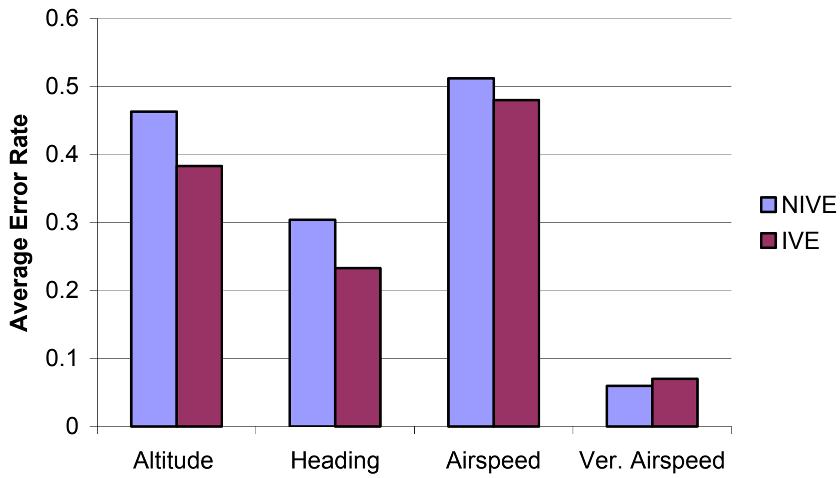


Figure 17: Average Error Rate Distribution for CSCD Tasks

# 7. DISCUSSIONS AND SUMMARY

## 7.1 Discussions

The general hypothesis tested was whether there is equal improvement in APC (measured by error rate) for people trained under IVE and NIVE. Three analyses were performed. First, we compared error rate performance for 1st and 5th trials across all tasks. Table 11 shows the result.

Table 11: Comparing NIVE and IVE for 1st & 5th Trial Performance

| | Altitude | Heading | Airspeed | Vert.Airspeed |
|---|---|---|---|---|
| NCAL | NIVE | No change | No change | No Change |
| GA | NIVE | NIVE | IVE | NIVE |
| CSCD | No change | No change | No change | No change |

As shown in Table 11, NIVE group shows performance gains in altitude error rate across all tasks, as well as improvement for heading and vertical airspeed for GA task. The IVE group showed performance gain over the NIVE group for airspeed error rate under GA task. There were no other observable gains especially for CSCD tasks and heading, airspeed, and vertical airspeed error rates under the NCAL task.

Second, we compared the last (5th) trial performance for both NIVE and IVE. The result is shown in Table 12. Here, NIVE and IVE groups did not show any statistically significant decrements in error rate for NCAL under heading and airspeed variables or for all CSCD task variables. However, the NIVE group showed decrements in altitude and airspeed error rates, while the IVE group showed decrements in heading and vertical airspeed error rates.

Table 12: Comparing NIVE and IVE for 5th Trial Performance

| | Altitude | Heading | Airspeed | Vert.Airspeed |
|---|---|---|---|---|
| NCAL | Same | Same | IVE | NIVE |
| GA | NIVE | IVE | NIVE | IVE |
| CSCD | Same | Same | Same | Same |

Lastly, we conducted an analysis to compare the overall error rate performance for NIVE and IVE groups (Table 13). The NIVE group performed better than the IVE with altitude error rates under NCAL and GA tasks. The IVE group performed better than the NIVE group for heading error rate under NCAL. There were no statistical significance differences using either the NIVE or IVE group for all tasks marked **"Same"** in the Table.

Table 13: Comparing NIVE and IVE for Overall Error Rate Performance

| | Altitude | Heading | Airspeed | Vert.Airspeed |
|---|---|---|---|---|
| NCAL | NIVE | IVE | Same | Same |
| GA | NIVE | Same | Same | Same |
| CSCD | Same | Same | Same | Same |

There are at least three observations to be drawn from the results of this experiment.

(1) The desktop (NIVE) training environment seems to show greater performance gains as measured by error rate compared to the immersive virtual environment. This observation confirms the finding of Ortiz (1994) in which performance of ab initio pilots were compared while flying a square pattern within specified performance limits with a PC-based simulator and a control group. The PC-based group was found to perform better than the control group. Although the experiment did not compare an immersive virtual environment, the findings can be used to validate performance results in similar settings.

(2) When both NIVE and IVE groups were compared on the error rate reductions over training trials, there were no noticeable differences in error rate decrement between them. However it seems that NIVE can be more useful in training control of vertical airspeed under the NCAL task, and control of altitude and airspeed under the GA task. On the other hand, IVE can be useful in training airspeed control under NCAL, and heading and vertical airspeed control under GA task. With the immersive environment, Patrick et al. (2000) noted that people are augmented with increased peripheral vision and capability to freely look around the surroundings. These capabilities are probably the reason the IVE group has performance gains during the NCAL task for airspeed control, and heading and vertical airspeed control for GA tasks.

(3) When error rates are compared across the number of trials and training environments, there are no exciting cost performance gains of IVE over NIVE. However, NIVE can provide performance gains for altitude control error rates for NCAL and GA tasks, while IVE can provide performance gains in heading control under NCAL tasks. This result, although derived from different test conditions, can be compared with the observations made by Hennessy, Wise, and Koonce (1995). In their study, they compared the

performance of pilots under traditional Instrument Landing System (ILS) and pathway-in-the-sky augmented display (a pseudo immersive VR system). The result showed that the PC-based (ILS group) provided a better measure of performance.

### 7.2 Summary

Overall, the results obtained from the current experiment do not justify any cost-saving advantage of IVE over NIVE. Because of the task specific gains in using either IVE or NIVE, we can reason that there are in fact some "opportunistic" cost savings based on these specific applications that lack generality across tasks and contexts. In addition, the fact that either IVE or NIVE provides increase in piloting task performance in some tasks needs to be considered in any training investment decision.

In previous research by Peterson, Wells, Furness, and Hunt (1998), maneuvering performance as measured by the precision with which the subject's ability to replicate a navigation route was experimentally tested in NIVE and IVE. The result was shown to be better for a nonimmersive VR environment (desktop with joystick) than the virtual motion controller (VMC). Similarly, in a study by Lampton et al. (1995), performance differences between a low-cost HMD (IVE condition) and standard PC-based simulator and monitor were evaluated using two groups for distance estimation tasks. The result showed that distance estimation was less accurate with the PC-based group. These result variations indicate that the cost tradeoff between the use of IVE over NIVE are task dependent and influenced by the fidelity of the training environments (Ortiz, 1994 ).

There are at least four factors that may contribute to the current results. These are:

(1) Some performance problems with HMD includes, but is not limited to field of view (FOV) and total field of regard (FOR) (Gallimore, Brannon, & Patterson, 1998). Limited FOV may have contributed to the poor performance of the IVE group because of the constrained range of display view.
(2) Display resolution (Naish & Miller, 1980). For example, display resolution for the HMD was more limited than the NIVE equivalent.
(3) Fatigue of the eyes may also have an effect on performance decrements of the IVE group.
(4) Lack of a head tracker may also be a factor. In general, studies show that eye trackers provide sensory information about the spatial location of objects and elaborates the visual details of objects to improve accuracy in navigation tasks (Bliss, Tidwell, & Guest, 1997; Hendrix & Barfield, 1997).

Assessment of pilot skill learning in NIVEs and IVEs could be continued by examining the experience of subjects' performance after the five trials to see if the two environments would differ any more or less than they already do. Use of subjective workload measures can be used to achieve this. Future work could also include trying the same study with various cockpits of other planes other than the Cessna airplane. Variations of noise distractions (weather, turbulence, etc.) during flight can also be assessed. Future studies should also investigate the effects of FOV, FOR, and display resolutions on performance. The effect of using a head tracker with an HMD may provide different results than the ones obtained here.

# 8. REFERENCES


Albert, W.S., Rensink, R.A., & Beusmans, J.M. (1999). Learning relative directions between landmarks in a desktop virtual environment. Spatial Cognition and Computation, 1, 131-144.

Amalberti, R., & Deblon, F. (1992). Cognitive modeling of fighter aircraft process control: A step towards an intelligent onboard assistance system. International Journal of Man-Machine Studies, 36, 639-671.

Anderson, J.R. (1993). Rules of the mind. Hillsdale, N.J: Lawrence Erlbaum and Associates.

Augusteijn, M.F., Broome, R.W., Kolbe, R.B., & Ewell, R.N. (1992). ITS challenger ---A domain-independent environment for the development of intelligent tutoring systems. Journal of Artificial Intelligence in Education, 392, 183-205.

Aukstakalnis, S., & Blatner, D. (1992). Silicon mirage: The art and science of virtual reality. Berkeley, CA: Peachpit.

Bell, H.H., & Waag, W.L. (1998). Evaluating the effectiveness of flight simulators for training combat skills: A review. The International Journal of Aviation Psychology 8, 223-242.

Bilodeau, E.A., & Bilodeau, M.I. (1961). Motor skill learning. Annual Psychology, 12, 243-280.

Bliss, J.P., Tidwell, P.D., & Guest, L. (1997). The effectiveness of virtual reality for administering spatial navigation training to firefighters. Presence, 6(1), 73-85.

Caro, P. (1988). Flight training and simulation. In E. Wiener & P. Nagel (Eds) Human Factors in Aviation (pp. 229-261). San Diego, CA: Academic Press.

Chambers, A.B. & Nagel, D.C. (1985). Pilots of the future: Human or computer? Communications of ACM, 28(11).

Chou, C.D., Madhavan, D., & Funk, K. (1996). Studies of cockpit task management errors. International Journal of Aviation Psychology, 6(4), 307-320

Connolly, T.J., Blackwell, B.B., & Lester, L. (1989). A simulator-based approach to training in aeronautical decision making. Aviation, Space, and Environment Medicine, 60(1), 50-52.

Cooke, N.J. (1994). Varieties of knowledge elicitation techniques. International Journal for Human-Computer Studies, 41, 801-849.

De Keysers, V. (1987). How can computer-based visual displays aid operators? International Journal of Human-Machine Studies, 27, 471-478.

Dennis, K.A. & Harris, D. (1998). Computer-based simulation as adjunct to an intro to flight training. The International Journal of Aviation Psychology, 8(3), 261-276.

Dion, D.P., Smith, B.A., & Dismukes, P. (1996). The cost/fidelity balance. Modern Simulation & Training: The International Training Journal, 2, 38-45.

Dix, A., Finlay, J. Abowd, G., & Beale, R. (1993). Human Computer Interaction. Prentice-Hall: New York.

Ericsson, K.A., & Simon, H.A. (1993). Protocol Analysis: Verbal Reports as Data (Revised Edition). London: MIT Press.

Fischer, E. (1979). The role of cognitive switching in head-up displays. (NASA Contractor report 3137). Moffett Field, CA: NASA Ames Research Center.

Fitts, P. M. (1956). Perceptual motor skill learning. In A.W. Melton (Ed.), Categories of Human Learning (pp. 27-33). New York: Academic Press.

Fitts, P.M., & Posner, M.I. (1967). Human Performance. Monterey, CA: Brooks and Cole Publications.

Gagne, R.E. (1962). The acquisition of knowledge. Psychological Review, 69, 355-365.

Galimore, J.J., Brannon, N.G., & Patterson, F.R. (1998). The effects of field-of-view on pilot head movement during low altitude flight. In Proceedings of the 42nd Annual Meeting of the Human Factors and Ergonomics Society(pp. 6-10). Santa Monica, CA: HFES.

Harper, S. D. (1991). Comparison of the training transfer of a personal computer-based simulator and a FAA-approved simulator in a general aviation aircraft. Dayton, OH: Wright State University.

Heally, A., Clowson, D., MacNamara, D., Marmie, W., Schneider, V., Rickard, T., Crutcher, R., King, C., Ericsson, K., & Broune, L. (1995). The long-term retention of knowledge and skills. The Psychology of Learning and Motivation, 30, 135-164.

Helmreich, R.L. (1984). Cockpit management attitudes. Human Factors, 26, 583-589.

Hendrix, C., & Barfield, W. (1997). Spatial discrimination in three-dimensional displays as a function of computer graphics eyepoint elevetaion and stereoscopic viewing. Human Factors, 39(4), 602-617.

Hennessy, L.W., III, Wise, J.A., & Koonce, J.M. (1995). The effects of a pathway-in-the-sky display on the performance of a two-axis tracking task by instrument-rated pilots. In R.S. Jensen & L.A. Rakovan (Eds.), Proceedings of the Eight International Symposium on Aviation Psychology (pp. 1073-1077). Columbus, OH: Ohio State University.

Hoc, J.M. (1993). Some dimensions of cognitive typology of process-control situation. Ergonomics, 36(11), 1445-1455.

Holyoak, K. J. & Thagard, P. (1996). Mental leaps: Analogy in creative thought. Cambridge, MA: The MIT Press.

Hirtle, S.C., & Hudson, J. (1991). Acquisition of spatial knowledge of routes. Journal of Experimental Psychology, 11, 335-345.

Hunt, G.I.F., & Hunt, L.M. (1986, December). Computer-based training. Interface, pp 50-51.

Jaeger, B.K. (1998). The effects of training and visual detail on accuracy of movement production in virtual and real-world environments. Proceedings of the Human Factors.and Ergonomics Society 42nd Annual Meeting (pp. 1486-1490).Santa Monica, CA: Human Factors & Ergonomics Society.

Johnson, H. & Johnson, P. (1991). Task knowledge structures: Psychological basis and integration into system design. Acta Psychologica, 78, 3-26.

Johnston, A.N., & Maurino, D. (1990). Human factors training for aviation personnel. ICAO Journal, 16-19.

Kelly, M.J. (1988). Performance measurement during simulated air-to-air combat. Human Factors, 30, 495-506.

Koonce, J. M. (1984). A brief history of aviation psychology. Human Factors, 26, 499-508.

Koonce, J.M., & Bramble, W. J. (1998). Personal computer-based flight training devices. The International Journal of Aviation Psychology, 8, 277-292.

Lampton, D.R., McDonald, D.P., Singer, M., & Bliss, J.P. (1995). Distance estimation in virtual environments. Proceedings of the Human Factors and Ergonomics Society 42nd Annual Meeting (pp. 1268-1272).Santa Monica, CA: Human Factors & Ergonomics Society.

Lintern, G. (1980). Transfer of landing skill after training with supplementary visual cues. Human Factors, 22 (1), 81-88.

Lintern, G., Thomley-Yates, K.E., Nelson, B.E., & Roscoe, S.N. (1987). Content, variety, and augmentation of simulated visual scenes for teaching air-to-ground attack. Human Factors, 31, 87-99.

MacFarlane, R. (1996). Simulation as an instructional procedure. In G. J.F. Hunt (Ed.). Designing instruction for human factors training in aviation (pp. 59-93).

Michalak, D.F. (1981). The neglected half of training. Training and Development Journal, 35, 22-28.

Niash, J.M., & Miller, D.L.(1980). An experimental evaluation of head-up display formats (NASA Technical Paper 1550). Moffett Field, CA: NASA Ames Research Center.

Ortiz, G.A. (1994). The effectiveness of PC-based flight simulation. The International Journal of Aviation Psychology,4, 285-291.

Patrick, E., Cosgrove, D., Slavkovic, A., Rode, J.A, Verratti, T. & Chiselko, G. (2000).Using a large projection screen as an alternative to head-mounted displays for virtual environments. Human factors in computing systems: CHI Conference proceedings. New York, NY: Association for Computing Machinery.

Pausch, R. (1993). Three views of virtual reality: An overview. Computer, 26 (6), 79-80.

Peterson, B., Wells, M., Furness, T.A., III, & Junt, E. (1998). The effects of the interface on navigation in virtual reality environments. Proceedings of the Human Factors.and Ergonomics Society 42nd Annual Meeting (pp. 1496-1500).Santa Monica, CA: Human Factors & Ergonomics Society

Piaget, J. (1970). The child's conception of movement and speed. Translated by G.E.T. Holloway and M.J. Mackenzie (Eds). London: Routledge and Kegan Paul Press

Piantania, T., Boman, D.K., & Gille, J. (1995). Human perceptual issues and virtual reality. Virtual Reality Systems, 43-52.

Pimental, K. & Teixera, K. (1995). Virtual reality: Through the new looking glass. New York: McGraw-Hill.

Povenmire, H.K., & Roscoe, S.N. (1971). An evaluation of ground-based flight trainers in routine primary flight training. Human Factors, 13, 109-116.

Presson, C.C., & Montello, D.R. (1994). Updating after rotational and translational body movements: Coordinate structure of perspective space. Perception, 23, 1447-1445.

Psotka, J. (1995). Immersive training systems: Virtual reality and education training. Instructional Sciences, 23, 405-431.

Rasmussen, J. (1986). Information processing and human-machine interaction: An approach to cognitive engineering. New York: North-Holland

Regian, J.W., & Shebilske, W. (1992). Virtual reality: An instructional medium for visual-spatial tasks. Journal of Communication, 42, 136-149.

Rich, E. (1978). Principles of Categorization. In Cognition & Categorization (E. Rosch &
B.B. Lloyd, Eds.). Hillsdale, NJ: Erlbaum.

Rich, E. (1983). Users are individuals: Individualizing user models. International Journal of Man-Machine Studies, 18, 199-214.

Rieser, J. J., Lockman, J.L., & Pick, Jr., H.L. (1980). The role of visual experience in knowledge of spatial layout. Perception & Psychophysics, 28, 185-190.

Rieser, J.J. (1989). Access to knowledge of spatial structure at novel points of observation. Journal of Experimental Psychology: Learning, Memory and Cognition, 15, 1157-1165.

Rose, F.D. & Attree, E.A. (2000). Training in virtual environments: Transfer to real-world tasks and equivalence to real-task training. Ergonomics, 43(4), 494-512.

Ruddle, R.A., Payne, S.J., & Jones, D.M. (1999). Navigating large-scale virtual environments: What differences occur between helmet-mounted and desktop displays? Presence: Teleoperators and Virtual Environments, 8(2), 157-168.

Seamester, T.L., Redding, R.E., & Kaempf, G.L. (1997). Applied cognitive task analysis in aviation. Brookfield, USA: Avebury Aviation.

Self, J.A. (1974). Student modes in computer-aided instruction. International Journal of Man-Machine Studies, 6(2), 261-276.

Skinner, B.F. (1953). Science and human behavior. New York: MacMilan Press.

Slater, M. & Usoh, M. (1993). Representation systems perceptual position and presence in immersive virtual environments. Presence, 2(3), 221-223.

Sternberg, R.J. (1999). Cognitive psychology. Fort Worth: Harcourt Brace College Publishers.

Stich, S. (1980). What every speaker recognizes. Behavioral and Brain Sciences, 3, 39-40.

Tannenbaum, S.I., Cannon-Bowers, J.A., Salas, E., & Mathieu, J.E. (1992). Deriving theoretically-based principles of training effectiveness to optimize training system design. Proceedings of the 14th Annual Interservice/Industry Training Systems Conference (pp. 619-631). San Antonio, TX: National Security Industrial Association.

Tefler, R. & Bent, J. (1992). Producing a workshop for training airline instructors. The Journal of Aviation/Aerospace Education and Research, 2 (2), 31-38.

Thorndike, P.W. & Hayes-Roth, B. (1982). Differences in spatial knowledge acquired from maps and navigation. Cognitive Psychology, 21, 145-148.

Vicente, K.H. (1999). Cognitive work analysis: Towards safe, productive, and healthy computer-based work. Mahwah, NJ: Lawrence Erlbaum Associates.

Williams, H.P., Hutchinson, S., & Wickens, C.D. (1996). A comparison of methods for geographic knowledge in simulated aircraft navigation. Human Factors, 38, 50-64.

Wenzel, E.M. (1992). Localization in virtual acoustic displays. Presence, 1, 80-107.

Wickens, C.D., Gordon, E., & Liu, Y.(1997). An introduction to human factors engineering. New York: Addison-Wesley.

Yeh, M. & Wickens, C.D. (1997). Effects of color-coding, intensity coding, and decluttering on visual search in electronic map displays (Technical report). Urbana, IL: University of Illinois Engineering Research Laboratory.